



Linked Data Applications: **There is no One-Size-Fits-All Formula**

Asunción Gómez-Pérez

Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid

<http://www.oeg-upm.net>

asun@fi.upm.es

Acknowledgements:

O. Corcho, D. Garijo, D. Vila, L. Vilches, B. Villazón

Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0

1. Introduction and Motivation
2. The process
3. Examples
 - Libraries: <http://datos.bne.es>
 - Geo: <http://geo.linkeddata.es/>
 - Metereology: <http://aemet.linkeddata.es/>
 - Travelling: <http://webenemasuno.linkeddata.es/>

Ontology Engineering Group

- Director: A. Gómez-Pérez
- Research Group (33 people)
- Participation in more than 15 EU projects, (3 as coordinator)
- Collaboration with many companies



A. Gomez-Perez



O. Corcho



G. Aguado



B. Villazon



Ontology Engineering Group Research Areas

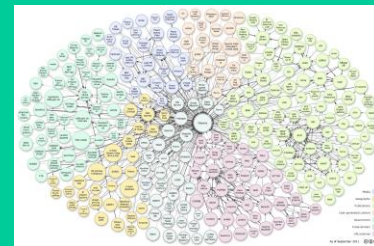
2004



Semantic e-Science
(Data Integration,
Semantic Grid)

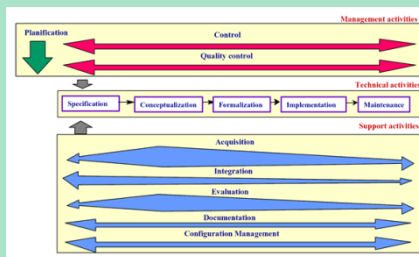
2009

Linked Data

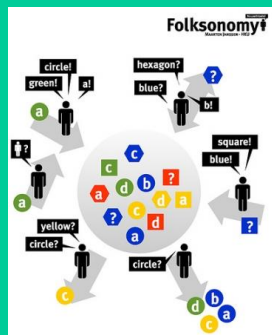


Ontological Engineering

1995



(Social)
Semantic Web



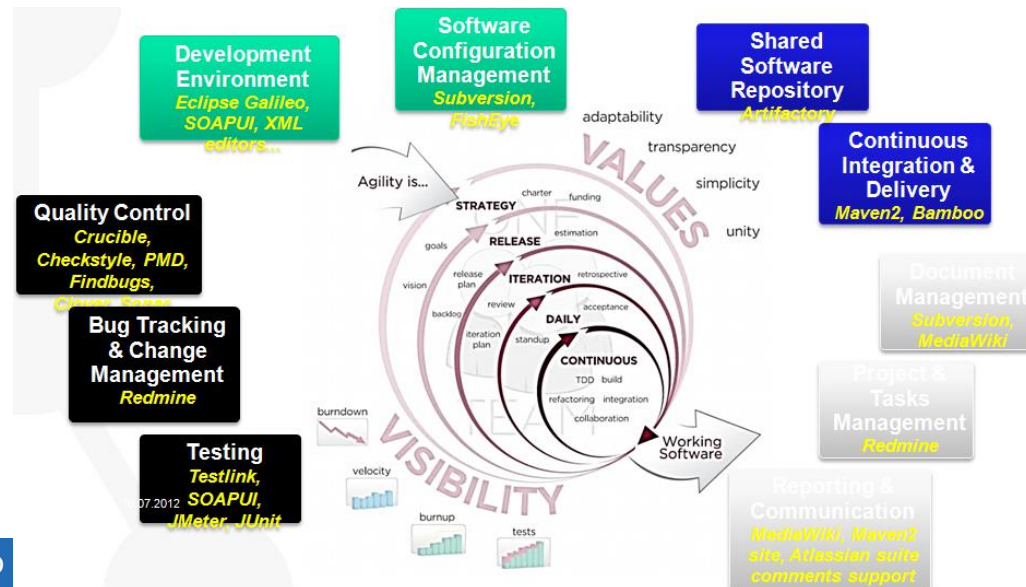
2000

Natural Language
Processing and
Multilingualism

1997



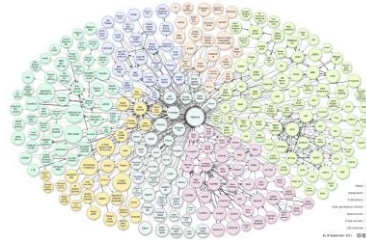
- Technology center funded by the Santander Group
 - Bank
 - Associated Software companies
- 1M€/year during the next five years
- Mission:
 - Open innovation ecosystem based on open software component developments
 - Managing open source software and products with LD



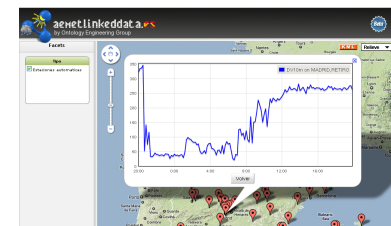
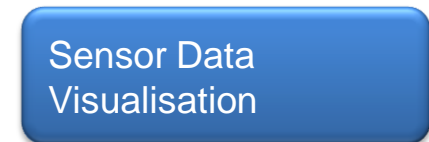
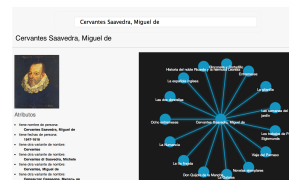
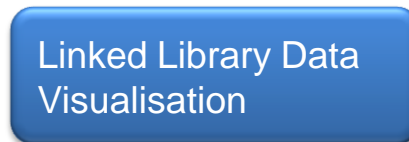
Linked data project at the Ontology Engineering Group



RDF Generation and Linking



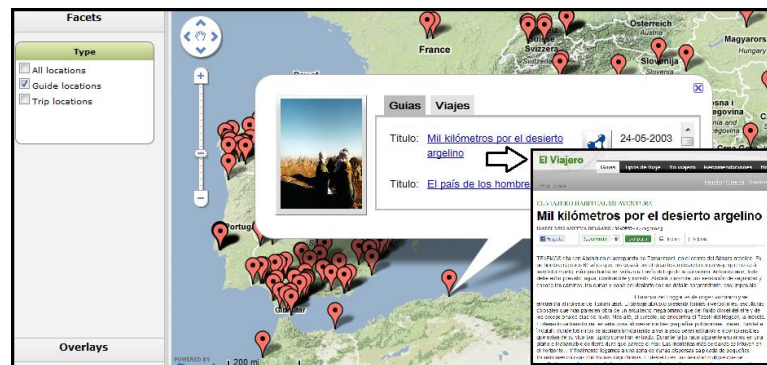
Visualization



Linked data: applications

Geo: <http://geo.linkeddata.es/>

Travelling: <http://webenemasuno.linkeddata.es/>



Libraries: <http://datos.bne.es>
<http://bne.linkeddata.es/>

Metereology: <http://aemet.linkeddata.es/>

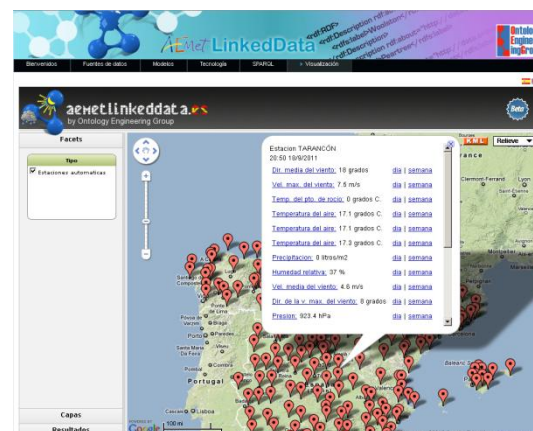
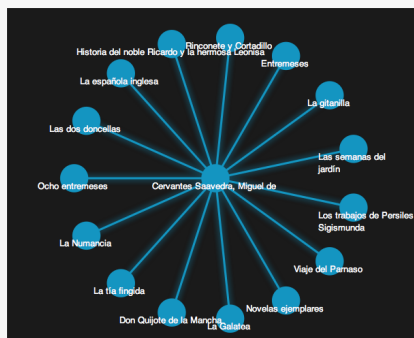
Cervantes Saavedra, Miguel de

Cervantes Saavedra, Miguel de

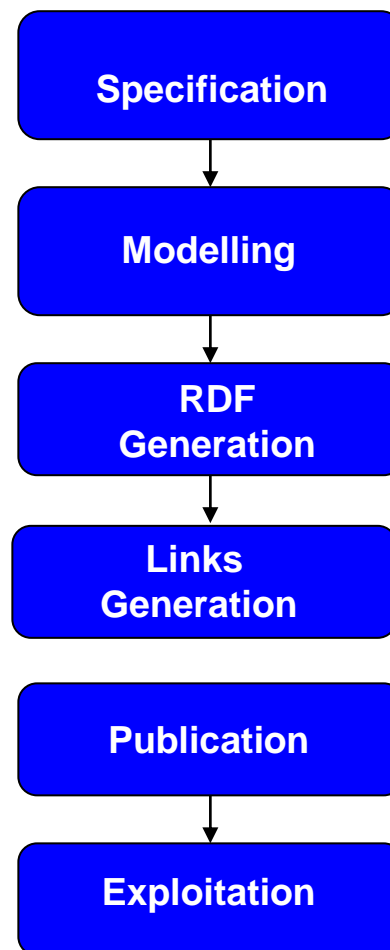


Atributos

- tiene nombre de persona: Cervantes Saavedra, Miguel de
- tiene fechas de persona: 1547-1616
- tiene otra variante de nombre: Cervantes
- tiene otra variante de nombre: Cervantes de Saavedra, Michele
- tiene otra variante de nombre: Cervantes, Miguel de
- tiene otra variante de nombre: Cervantes Saavedra, Mirel de



1. The concept
2. The process
3. Examples



1. The concept
2. The process
3. Examples



- Libraries: <http://datos.bne.es/>
<http://bne.linkeddata.es/>
- Geo: <http://geo.linkeddata.es/>
- Meteorology: <http://aemet.linkeddata.es/>
- Travelling: <http://webenemasuno.linkeddata.es/>

```
graph TD; A[Specification] --> B[Modelling]; B --> C[RDF Generation]; C --> D[Links Generation]; D --> E[Publication]; E --> F[Exploitation];
```

Specification

Modelling

RDF Generation

Links Generation

Publication

Exploitation

- Different communication formats:
 - **MARC 21 format for Bibliographic Data**
 - **MARC 21 format for Authority Data**
 - Others: Holdings, Classification, etc.
- Three main elements:
 - **Record structure**: ISO 2709. Fields, indicators, subfields...
 - **Content designation**: "Meaning" of codes and conventions
 - **Content**: Defined outside the MARC standard (ISBD, AACR..)

So, RDBtoRDF technologies were not appropriate for this task.

- Records in the **MARC 21** format
- **3.9 million** bibliographical records
- **4.2 million** authority records
- Version: November, 2011

Specification

Modelling

RDF Generation

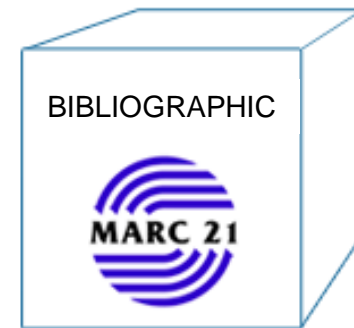
Links Generation

Publication

Exploitation



2492684	Persons
347737	Corporate bodies
84381	Conferences
715131	Titles
347793	Subject



76576	Maps
320727	Sound recordings
166017	Gravings, drawings, pictures
35770	Manuscripts
143959	Ancient books
2696560	Modern books
178473	Scores
3021	Electronic resources
156634	Serials
96672	Videos

Specification

- Authority record: Camus, Albert*

Control Field

001 XX1721208
 005 200012181124
 008 901120nn aijnnaabn n aaa
 016 \$a BNE19900178994
 040 \$a SpMaBN \$b spa \$c SpMaBN \$e rdc \$f
 embne

Field	Subfield	Content
100	10 \$a	Camus, Albert
	\$d	1913-1960

**HEADING
1XX**

670 \$a El mite de Sísif, 1987 \$b port. (Albert Camus)

670 \$a Dic. de filosofía, de J. Ferrater Mora, 1980 \$b (Camus., Albert (1913-1960); n. Mondovi, Argel)

670 \$a Aut. BN-OPALE, 1995 \$b (Camus, Albert)

* <http://datos.bne.es/resource/XX1721208>

MARC21 record content designation

- Authority record: Camus, Albert*

Control Number	001	XX1721208	HEADING – Personal	
Personal name	Name	100	10 \$a	Camus, Albert
Dates associated with name			\$d	1913-1960
Source consulted	Citation	670	\$a	El mite de Sísif, 1987 \$b port. (Albert Camus)

- Human reading:

*An authority record that describes **a Person**, named **Camus, Albert** with associated **dates 1913-1960***

* <http://datos.bne.es/resource/XX1721208>

Frequency of codes in records

[illegible]

Specification



Modelling



RDF Generation



Links Generation



Publication



Exploitation

- **Source data:** MARC 21 records, not RDB. Very flat structure difficult to map to richer models



Domain experts (catalogers) need to be part of the mapping process.

- Highly specialized library models: FRBR, ISBD.



Data quality good but still many errors: **data curation during the LD generation process**

- Iterative and incremental transformation process: **measure coverage and progress.**

- **Multilinguality**, collaboration with IFLA

Specification

Modelling

RDF Generation

Links Generation

Publication

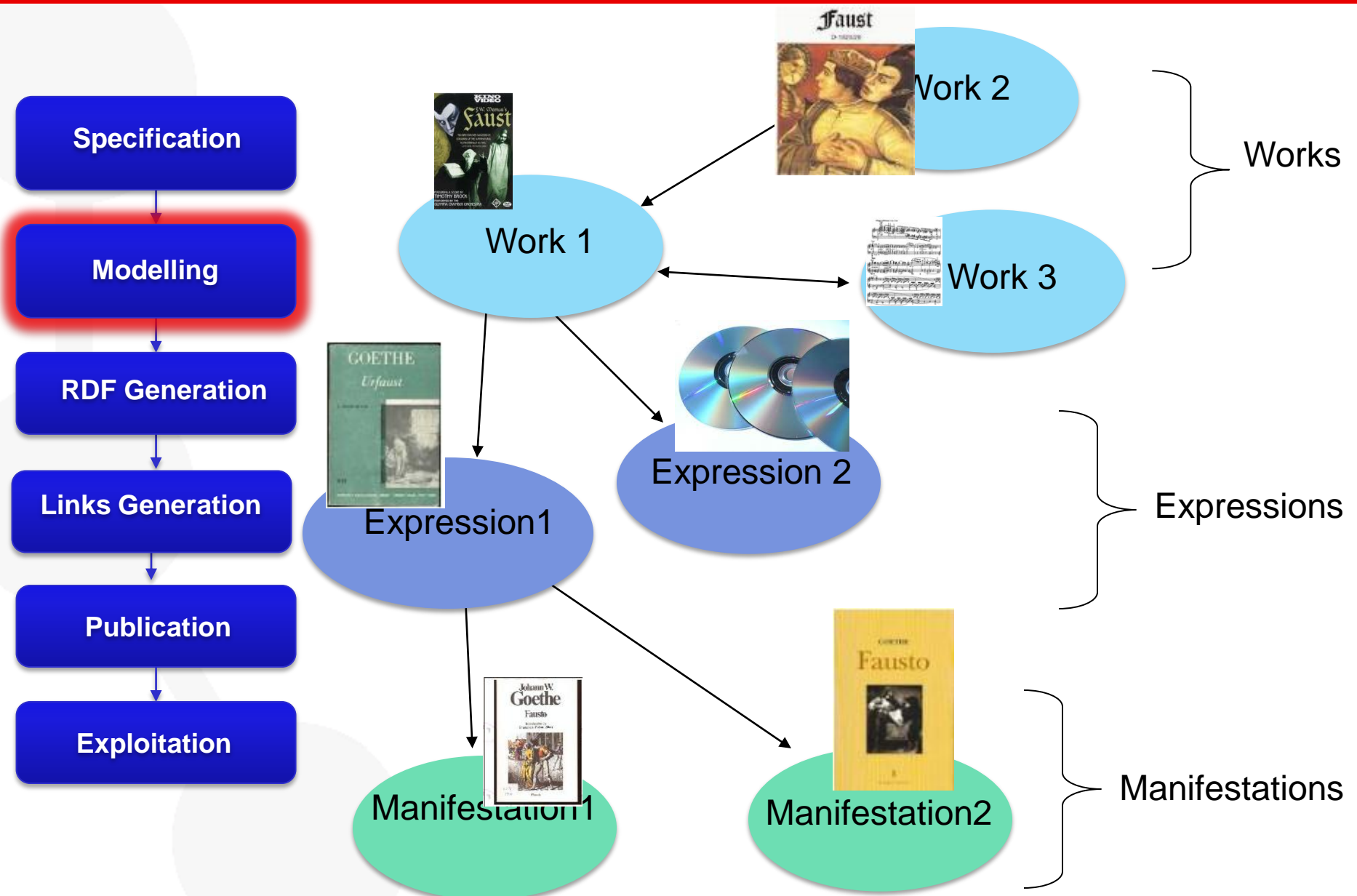
Exploitation

Reuse!

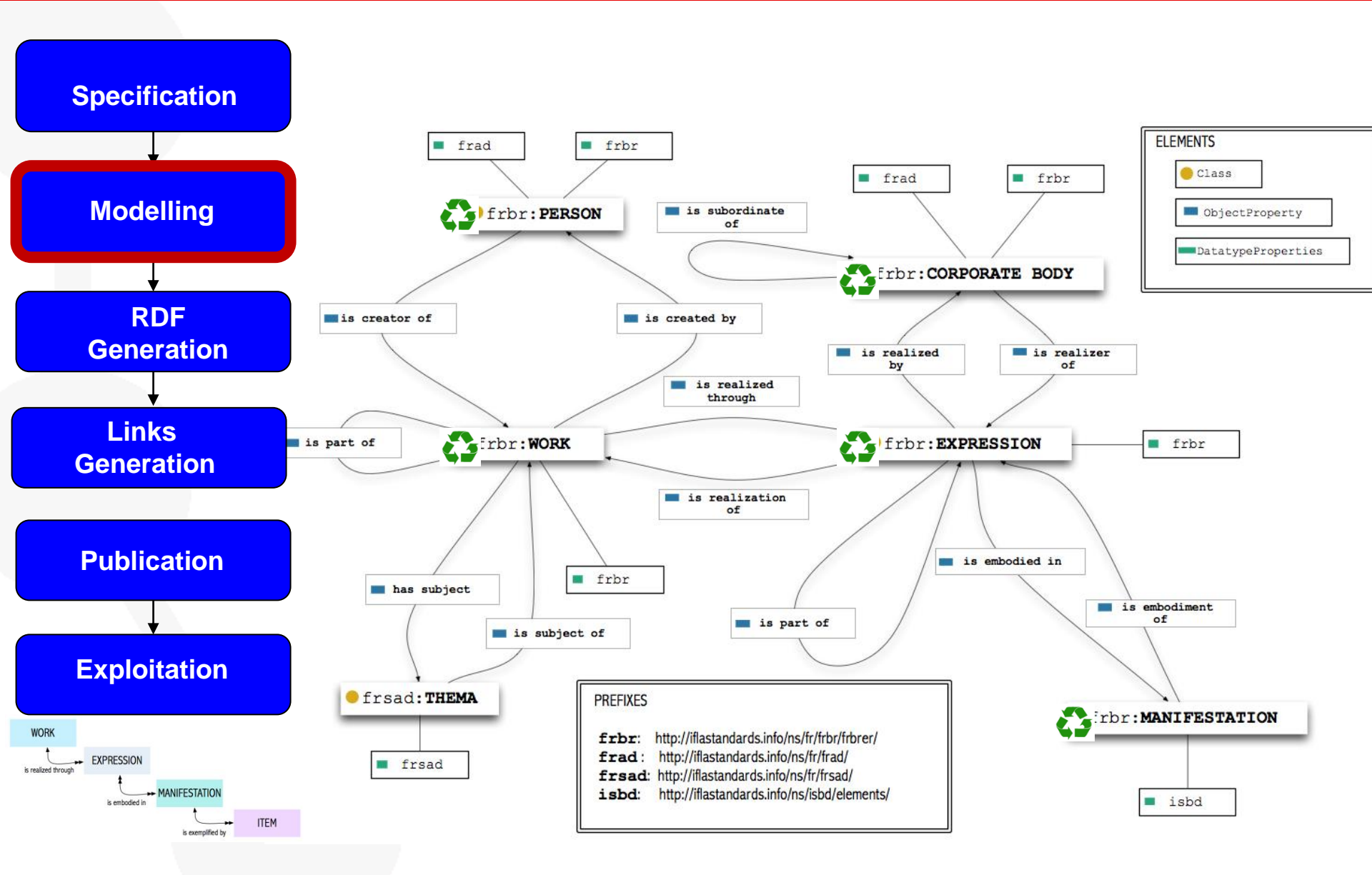
STANDARDS

Shared
Understanding





The Ontology: based on IFLA vocabularies



Who will be the mapping generator?

Specification

Modelling

RDF
Generation

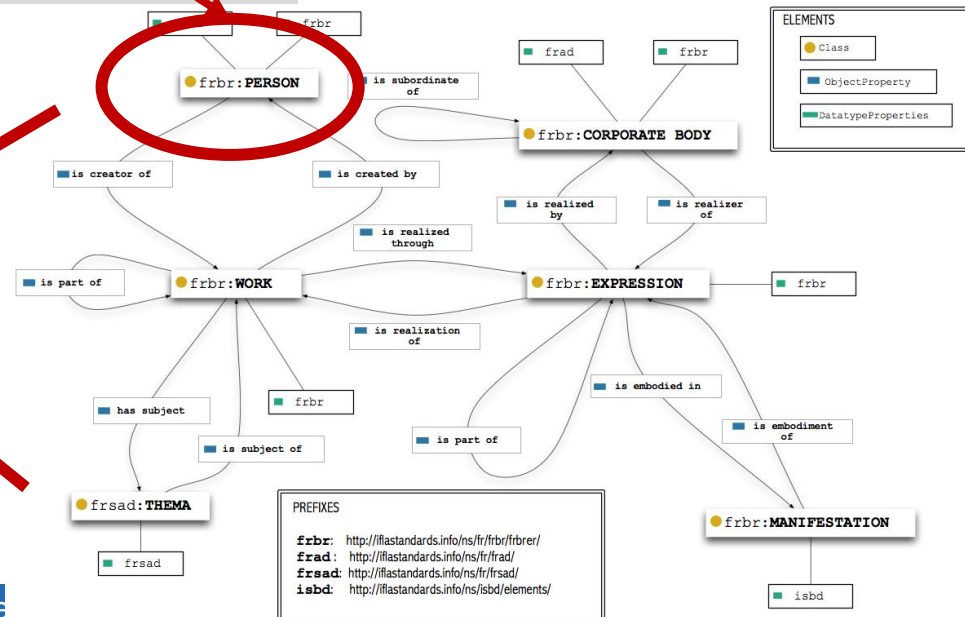
Links
Generation

Publication

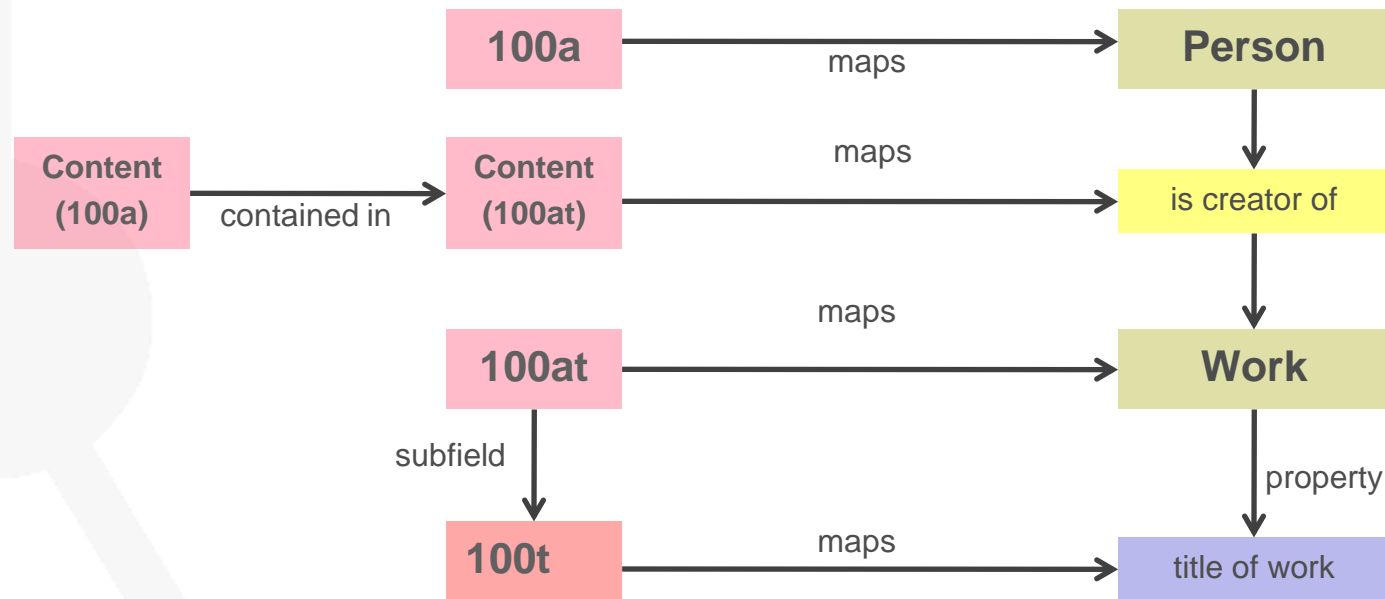
Exploitation



001 XX1721208
005 200012181124
008 901120nn aijnnaabn n aaa
016 \$a BNE19900178994
040 \$a SpMaBN \$b spa \$c SpMaBN \$e rdc \$f embne
100 10 \$a Camus, Albert
\$d 1913-1960
670 \$a El mite de Sísif, 1987 \$b port. (Albert Camus)
670 \$a Dic. de filosofía, de J. Ferrater Mora,
1980\$b(Camus., Albert (1913-1960); n. Mondovi,
Argel)
670 \$a Aut. BN-OPALE, 1995 \$b (Camus, Albert)



Similar to mapping ontologies



Marimba allows librarians to create mappings between MARC21 records and IFLA vocabularies using spreadsheets



Basic structure

MARC21 info	Records count	Content sample	Mapping
100 \$a \$d	888.880	Camus, Albert 1913-1960	foaf:Person
100 \$a	999.999	Cervantes, Miguel de	foaf:name
100 \$a \$m	10.000	Cervantes, iguel	ERROR



Classification
mapping



Annotation
mapping



Relationships
mapping

Librarians create mappings using excell

Classification mapping



	A	B	C	
1	Combination	Records count	Example	Mapping value
2	a	881115	[\$aYcardo, Maruja]	Person
3	at	685861	[\$aGual, Adrià, \$t Donzell qui cerca muller]	Work
4	adt	347783	[\$aVargas, Pedro, \$d1906-1989, \$tCanciones de nuestra vida]	Work
5	ad	183377	[\$aCornette de Saint Cyr, Pierre, \$d1939-]	Person
6	atl	87828	[\$aJames, Samantha, \$tSterling family trilogy, \$lEspañol (Barcelona)]	Expression
7	adtl	60231	[\$aSalvatore, R. A., \$d1959-, \$tThe sellswords, \$lEspañol (Barcelona)]	Expression
8	ae	35936	[\$aAbad, Atanasio, \$eimp.]	Person
9	ac	25992	[\$aA. M. B., \$cdottor]	Person
10	adv	19807	[\$aCavanna, Xavier, \$d1961-, \$vExposiciones]	Thema
11	act	17359	[\$aCoelho, J. R., \$c(m'xico), \$tBasta amor meu terno peito]	Work
12	ade	14573	[\$aMontoliu, Tete, \$d1933-1997, \$eint.]	Person
13	av	11427	[\$aBoisvert, Georges, \$vHomenajes]	Thema
14	aq	10143	[\$aValdÈs, Manuel, \$qValdÈs García]	Person
15	adx	8326	[\$aFunes, Gregorio, \$d1749-1829, \$xBiografías]	Thema
16	adtp	8167	[\$aBenevoli, Orazio, \$d1605-1672, \$tDeus pro nobis, \$pSanctus]	Work
17	ax	6563	[\$aGarret, Almeida, \$xCrítica e interpretaciOn]	Thema
18	acd	6440	[\$aAbad, Bautista, \$c(O.S.A.), \$d -ca. 1701]	Person
19	acdt	6262	[\$aTom's Moro., \$cSanto, \$d1478-1535, \$tThe last things]	Work
20	aqt	6133	[\$aIvarez, Pilar, \$qIvarez Saenz, \$tConvivir]	Work
21	adto	3480	[\$aTorres Nin, Lorenzo, \$d1890-1964, \$tLas pobres millonarias, \$oarr.]	Expression
22	adtmnr	3340	[\$aLeybach, Ignace Xavier Joseph, \$d1817-1891, \$tNocturnos., \$mpiano, \$r]	Work
23	adtk	3085	[\$aChaikovskiÛ, Piotr Ilich, \$d1840-1893, \$tEl lago de los cisnes (suite), \$k]	Expression
24	adtpo	2964	[\$aDauphin, LÉopold, \$d1847-1925, \$tDjanika, \$pNoble Époux, \$oarr.]	Expression
25	ace	2479	[\$aAcosta, Felipe de, \$c(O.P.), \$eed. lit]	Person
26	adtmn	2360	[\$aLigeti, György, \$d1923-2006, \$tCuartetos, \$mcuerda, \$nn. 1]	Work
27	adtn	2330	[\$aBernaola, Carmelo, \$d1929-2002, \$tSinfonías, \$nn. 3]	Work
28	adtnp	2220	[\$aBach, Johann Sebastian, \$d1685-1750, \$tMatthæuspassion., \$nN. 20., \$r]	Work
29	adxv	2155	[\$aFiguera Aymerich, Ingela, \$d1902-1984, \$xHomenajes, \$vVideograbacio]	Thema
30	adtm	1990	[\$aGorriti, Felipe, \$d1839-1896, \$tMagnificat, \$m4 voces, orquesta]	Work

Librarians create mappings using excell

	Annotation mapping			
	A	B	C	D
1	Combinatio	Records c	Example	Mapping value
2	260b	3892943	Imp. Lemerrier et C.ie	http://iflastandards.info/ns/isbd/elements/P1117
3	245a	3871272	Polka elegante para piano	http://iflastandards.info/ns/isbd/elements/P1004
4	040a	3870286	SpMaBN	
5	260a	3845589	[Paris	http://iflastandards.info/ns/isbd/elements/P1022
6	300a	3827202	2 p.	http://iflastandards.info/ns/isbd/elements/P1022
7	245h	3751432	M'sica impresa	http://iflastandards.info/ns/isbd/elements/P1185
8	260c	3714336	1865-1869]	http://iflastandards.info/ns/isbd/elements/P1018
9	300c	3526395	22 cm	http://iflastandards.info/ns/isbd/elements/P1018
10	016a	3103593	bipaBNE20070006028	
11	245c	2805186	por Henri Cramer	http://iflastandards.info/ns/isbd/elements/P1007
12	017a	2589626	GC 30-2006	http://iflastandards.info/ns/isbd/elements/P1092
13	017b	2582999	Oficina Depósito Legal M¿laga	
14	6502	1707468	embne	
15	300b	1627566	aguafuerte	http://iflastandards.info/ns/isbd/elements/P1023
16	020a	1527638	84-7568-355-X	http://iflastandards.info/ns/isbd/elements/P1154
17	016a	1461610	IND9200366795	



	A	B	C	D
1	Variation	Sample	Relationship	
2	n	Sociedad Didáctico-Musical Método especial de solfeo cor	frbr:P2057 Is part of work frbr:P2058	
3	o	Bach, Johann Sebastian 1685-1750 Partita laúd		
4	p	Iglesia Católica Synodus Episcoporum (2008) Verbum Dor	frbr:P2057 ; frbr:P2058	
5	q	Isabel María Isabel María López Campos The face		
6	r	Chopin, Fryderyk 1810-1849 Valses piano mi menor		
7	s	Beethoven, Ludwig van 1770-1827 Fantasien pia		
8	t	Castellà Andreu, Josep M. La función constituci		

Relationships
mapping

Marimba interprets the Mappings and generate the RDF

Specification

```
001 XX1721208
```

```
.....
```

```
100 10 $a Camus, Albert
```

```
$d 1913-1960
```

```
.....
```



Modelling

- **Classify:** Exploiting the heading field and subfield **codes**.

100 \$a \$d → Person (it has a personal name)

100 \$a \$d \$t → Work (it has a title)

RDF
Generation

- **Annotate:** Using subfield codes and the **content**.

100 \$a "Camus, Albert" → frbr:3001 "Camus, Albert"

100 \$t "La Peste" → frbr:P3039 "La Peste"

Links
Generation

Publication

Exploitation



MARC 21 record (Input)	Action	RDF (Output)
100 \$a \$d	Classify	rdf:type frbr:C1005
100 \$a Camus, Albert	Annotate	frbr:P3039 "Camus, Albert"
100 \$d 1913-1960	Annotate	frbr:P3040 "1913-1960"

Mapping process more in detail

↓
**RDF
Generation**

- **But**, what about the **relationships** between the entities?
- Relationships between records are **not explicit** in MARC.

Goal: The work "La Peste" was created by Albert Camus

001 XX1721208

100 10 \$a Camus, Albert \$d 1913-1960

Common

Person

001 XX1910518

100 10 \$a Camus, Albert \$d 1913-1960 \$tLa peste

Common

Diff

Work

We know the type of R1 and R2, and we look at the heading diff

bne:XX1721208

frbr:2010

bne:XX1910518

(isCreatorOf)

* <http://datos.bne.es/resource/XX1910518>

Marimba: Mapping process summary

Specification

Modelling

RDF
Generation

Links
Generation

Publication

Exploitation



(MARC records)

001 XX1721208

100 10 \$a Camus, Albert \$d 1913-1960

Classify

bne:XX1721208 a **frbr:Person**

Annotate

bne:XX1721208 a frbr:Person
frbr:name "Camus, Albert" .
frbr:hasDates 1913-1960

Relate

bne:XX1721208 a frbr:Person
frbr:name "Camus, Albert" .
frbr:hasDates 1913-1960 .
frbr:isCreatorOf bne:XX1910518



001 XX1910518

100 10 \$a Camus, Albert\$d1913-1960 \$tLa
peste

bne:XX1910518 a **frbr:Work**

bne:XX1910518 a frbr:Work
frbr:title "La Peste"

bne:XX1910518 a frbr:Work
frbr:title "La Peste" .
frbr:isCreatedBy bne:XX1721208

Marimba uses the ontology to generate RDF



Specification

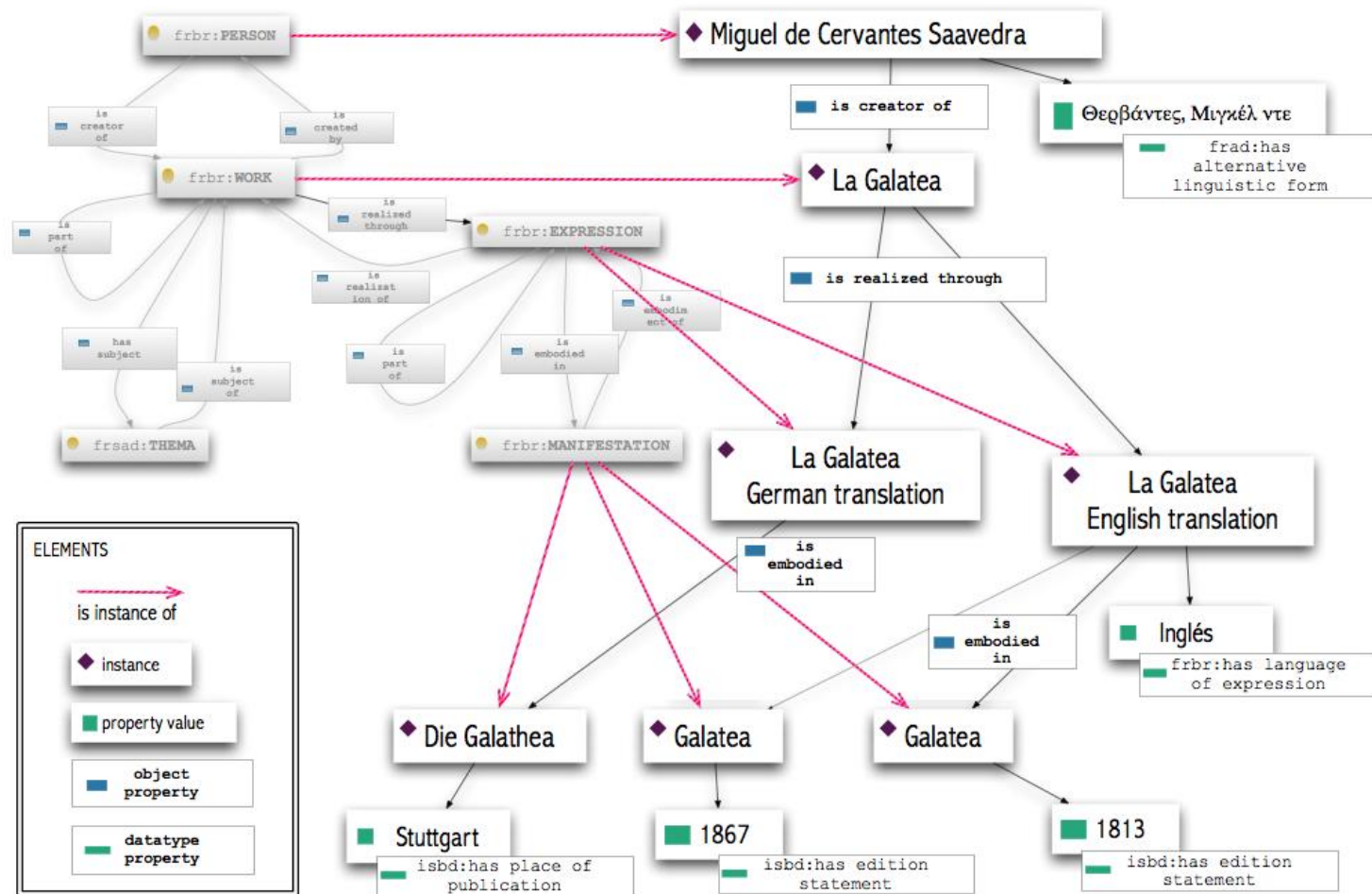
Modelling

RDF
Generation

Links
Generation

Publication

Exploitation



Marimba links with other resources: VIAF, DNB, SUDOC, LIBRIS, DBpedia

Specification

Modelling

RDF
Generation

Links
Generation

Publication

Exploitation

<http://d-nb.info/gnd/11851993X>

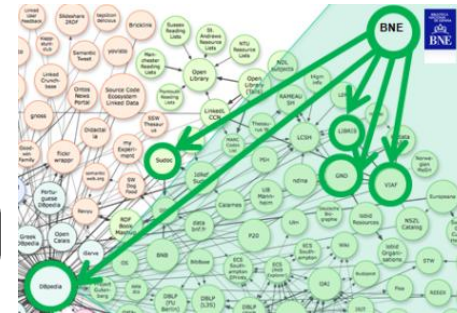


DNB

<http://viaf.org/viaf/17220427>



VIAF



Same As

Same As

Same As

http://dbpedia.org/resource/Miguel_de_Cervantes



DBpedia

<http://datos.bne.es/resource/XX1718747>



BNE

Same As

<http://www.idref.fr/026774771/id>



SUDOC

Same As

<http://libris.kb.se/resource/auth/45369>



LIBRIS

Specification

Modelling

RDF
Generation

Links
Generation

Publication

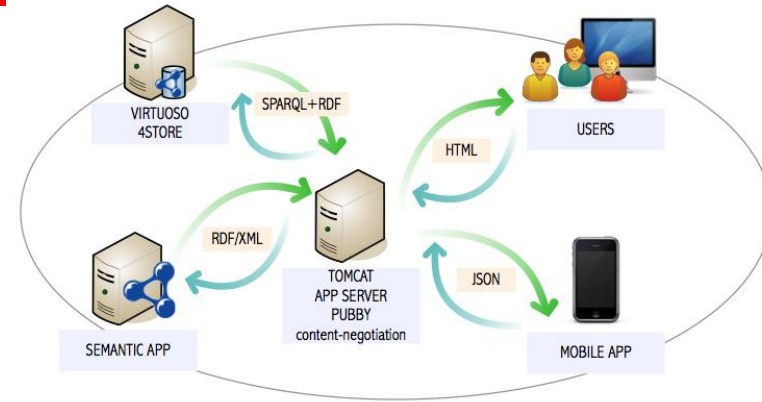
Exploitation

Data publication

Metadata publication using VOID

To facilitate the discovery

- Register in CKAN your dataset
- Use to **sitemap4rdf** to generate the site map
- Upload the site map to **Google** and **Sindice**



Specification

Modelling

RDF
Generation

Links
Generation

Publication

Exploitation

<http://bne.linkeddata.es/>

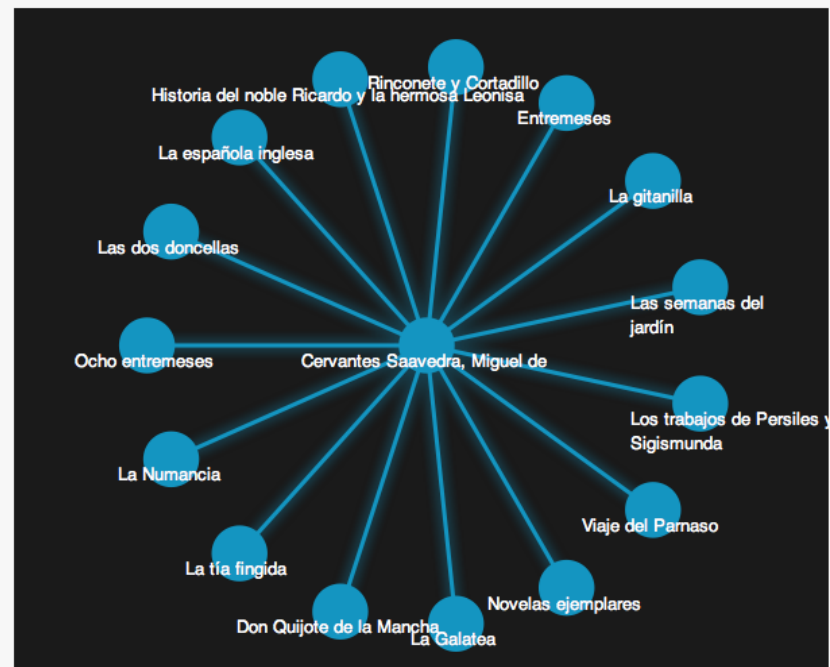
Cervantes Saavedra, Miguel de

Cervantes Saavedra, Miguel de



Atributos

- tiene nombre de persona:
Cervantes Saavedra, Miguel de
- tiene fechas de persona:
1547-1616
- tiene otra variante de nombre:
Cervantes
- tiene otra variante de nombre:
Cervantes di Saavedra, Michele
- tiene otra variante de nombre:
Cervantes, Miguel de
- tiene otra variante de nombre:
Сервантес Сааведра, Мигель де
- tiene otra variante de nombre:



- **Modelling: based on IFLA Vocabularies**

- Open Metadata Registry
- Neon Toolkit



- **Mapping and generation**

- **MARiMbA:** Library-oriented, supports and facilitates the entire process of transformation from MARC21 to RDF

- **Publication:**

- Virtuoso Universal Server
- Pubby
- CKAN registry
- Sitemap4rdf

```
<priority>1.00</priority>
<changefreq>daily</changefreq>
</url>
<url>
<loc>http://www.virtuopet1.net/taj/comento.php</loc>
<priority>0.80</priority>
<changefreq>daily</changefreq>
</url>
</robots>

< SITEMAP XML />

<url>
<loc>http://www.virtuopet1.net/taj/screenshots.php</loc>
<priority>0.80</priority>
<changefreq>daily</changefreq>
</url>
<url>
<loc>http://www.virtuopet1.net/taj/download.php</loc>
<priority>0.80</priority>
<changefreq>daily</changefreq>
</url>
</url>
```



- **Exploitation:**

- Web Applications that visualize data using SPARQL



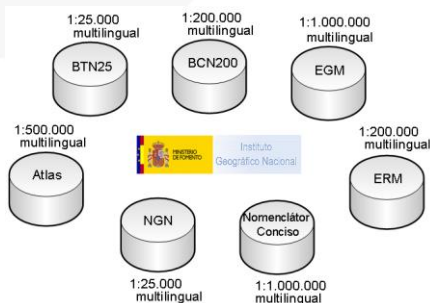
1. The concept
2. Foundations
3. The process
4. Examples



- Libraries: <http://datos.bne.es>
- <http://linkeddata3.dia.fi.upm.es/bne-demo>
- Geo: <http://geo.linkeddata.es/>
- Meteorology: <http://aemet.linkeddata.es/>
- Travelling: <http://webenemasuno.linkeddata.es/>

Specification

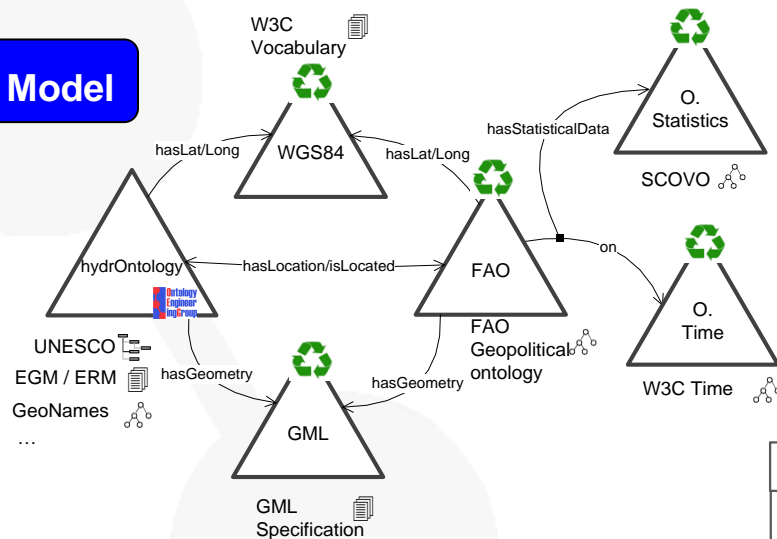
Uniform access to the Spanish Geographical Institute Databases



7 geographical DB

- Granularity
- Scale
- Multilingual

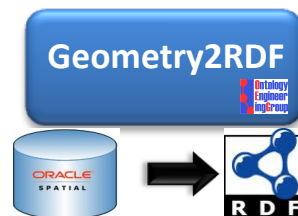
Model



Legend

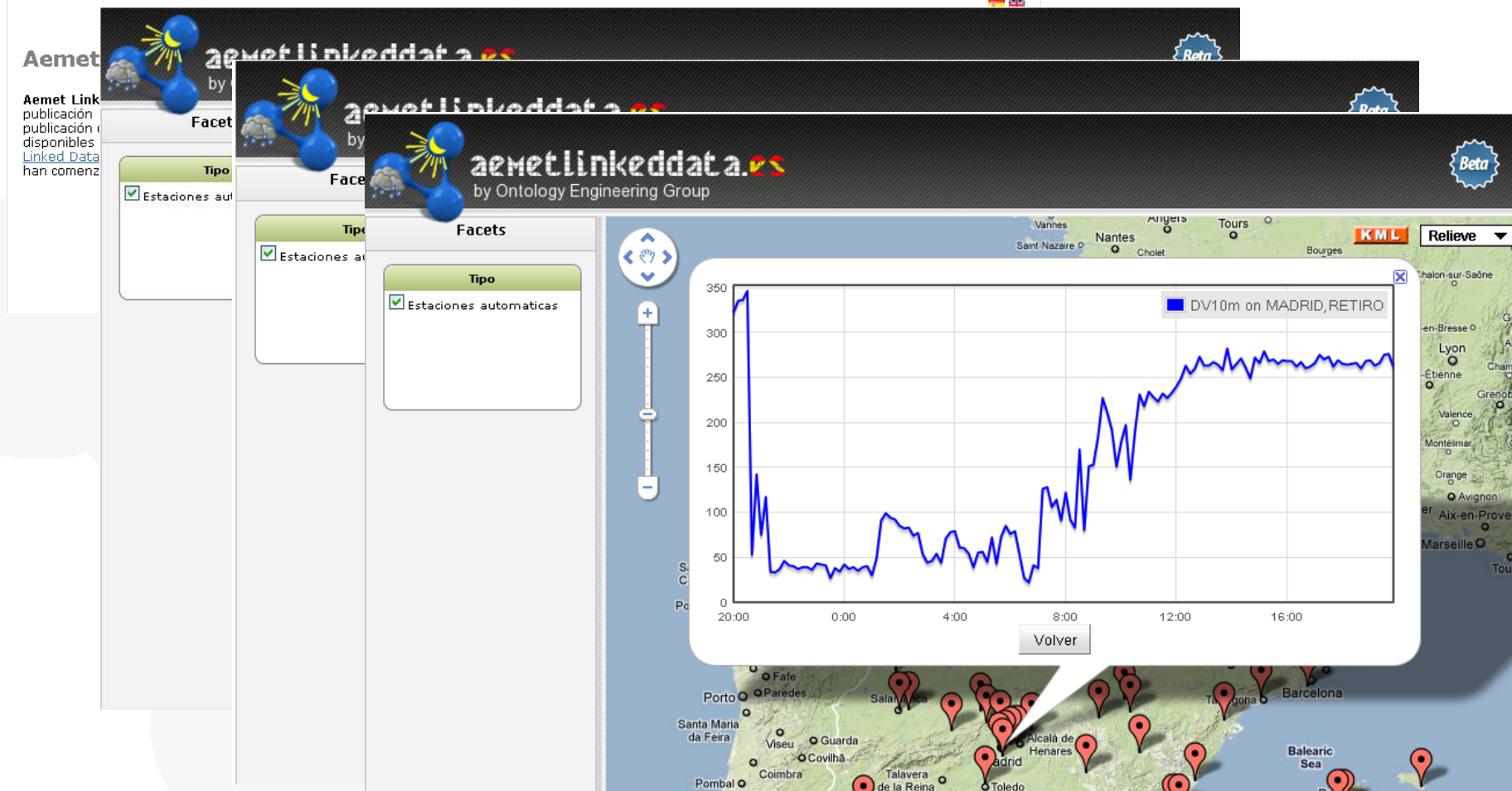
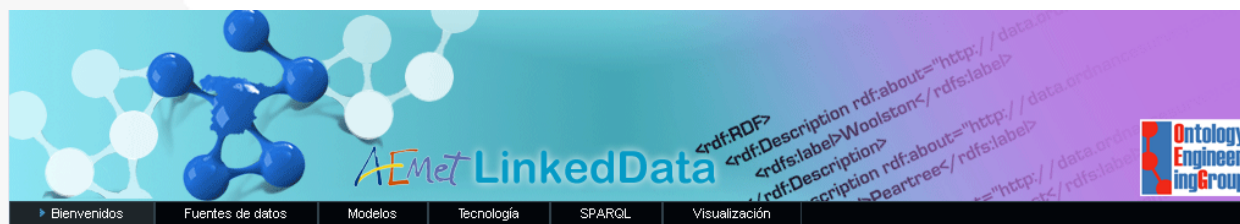
Ontology	
Specification	
Thesaurus	

RDF Generation from DB



shp2RDF







El Viajero

Facets

Capas

Filtrar por fecha

Linea temporal

Resultados

El Viajero

Introduce el nombre de un autor para comprobar su información relacionada

Búsqueda de autores

JOSÉ_CARLOS_CAPEL

Mostrando resultados 1-10 de 457











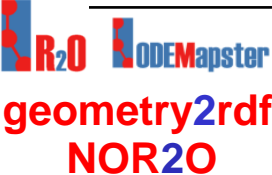










Anterior--Siguiente

El Viajero	Reconfortantes recetas italianas	Fecha de publicación: 02-04-2005
	Tipo de recurso: Guia	Creador: http://webenemasuno.linkeddata.es/elviajero/resource/Account/ACCJOS%C3%89_CARLOS_CAPEL
		Subtítulo: FORTUNATO, la última novedad madrileña dentro del bar de copas Fortuny
		Longitud: 641
		Medida de longitud: word
		Referencia en El País: http://elviajero.elpais.com/articulo/20050402elpvialbv_4/Tes
		La guía habla del lugar: http://geo.linkeddata.es/resource/ComunidadAut%C3%B3noma/Comunidad%20de%20Madrid
El Viajero	Más de doce sabores suculentos	Fecha de publicación: 14-08-2004
	Tipo de recurso: Guia	Creador: http://webenemasuno.linkeddata.es/elviajero/resource/Account/ACCJOS%C3%89_CARLOS_CAPEL

cional en la medina musulmana
uz del mediodía y enciende el
a, en la primera planta. O las
jlesia con apuntes mobiliarios de
nos espacios algo apretados por
e oyen los pasos, las
scasa para la lectura, y los
le todo reside en las plantas
mentos actuales. Si el salón de
literalmente sobre el abismo,
, la alberca también colgada

<DBY13046349521410>
<DO>

15>

Phase/Domain	Library (BNE)	Goegraphic (IGN, Otalex)	Meteorology (AEMET)	Travelling (PRISA)	Statistic (INE)
Modeling	  PROV DC	  Wgs84 time W3C	  SSN ontology PROV	 SIOC PROV	 Scovo Data cube
RDF generation	 	 R2O ODEMapster geometry2rdf NOR2O	 CSV parser	 CSV parser	 NOR2O
Links generation	 DNB VIAF LIBRIS DBPEDIA	Silk DBPEDIA Geonames	Silk Geolinkeddata.es	Silk DBPEDIA Geolinkeddata.es	NOR2O Geolinkeddata.es
Publication	   				
Exploitation	 map4rdf		SPARQL		

<http://datos.bne.es>

- Total number of authority records: **4.100.000**
- Total number of bibliographic records: **2.390.140**
- **Total number of RDF triples: 58.053.215**
- **Links (15% authority): 587.520**
- Linked sources:
 - VIAF
 - SUDOC (Sistema Universitario de Documentación) FR
 - GND (Auth German National Library) GER
 - LIBRIS Sweden
 - DBPedia

<http://geo.linkeddata.es/>

Number of geo type phenomenon: 95 (Rivers, mountains, etc.)

Number of geo entities: 155.000

Total number of RDF triples: 21.564.199

Links: 1002 (outlinks) y 6782 (coming)

Linked sources: DBpedia y GeoNames (outlinks)
AEMET y El Viajero (entry)

<http://webenemasuno.linkeddata.es/>

- Total number of guides: 27.876
- Total number of posts: 32.502
- Total number of locations: 6.838
- **Total number of RDF triples: 9.462.339**
- **Linked sources: 12.750**
 - DBPedia (6024 links)
 - GeoLinkedData (6726 links)





Linked Data Applications: **There is no One-Size-Fits-All Formula**

Asunción Gómez-Pérez

Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid

<http://www.oeg-upm.net>

asun@fi.upm.es

Acknowledgements:

O. Corcho, D. Garijo, D. Vila, L. Vilches, B. Villazón

Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0